Chiplet-Package Co-Design For 2.5D Systems Using Standard ASIC CAD Tools

MD Arafat Kabir, Yarui Peng
Computer Science and Computer Engineering Department
University of Arkansas, Fayetteville, AR, US

UNIVERSITY OF ARKANSAS

Introduction

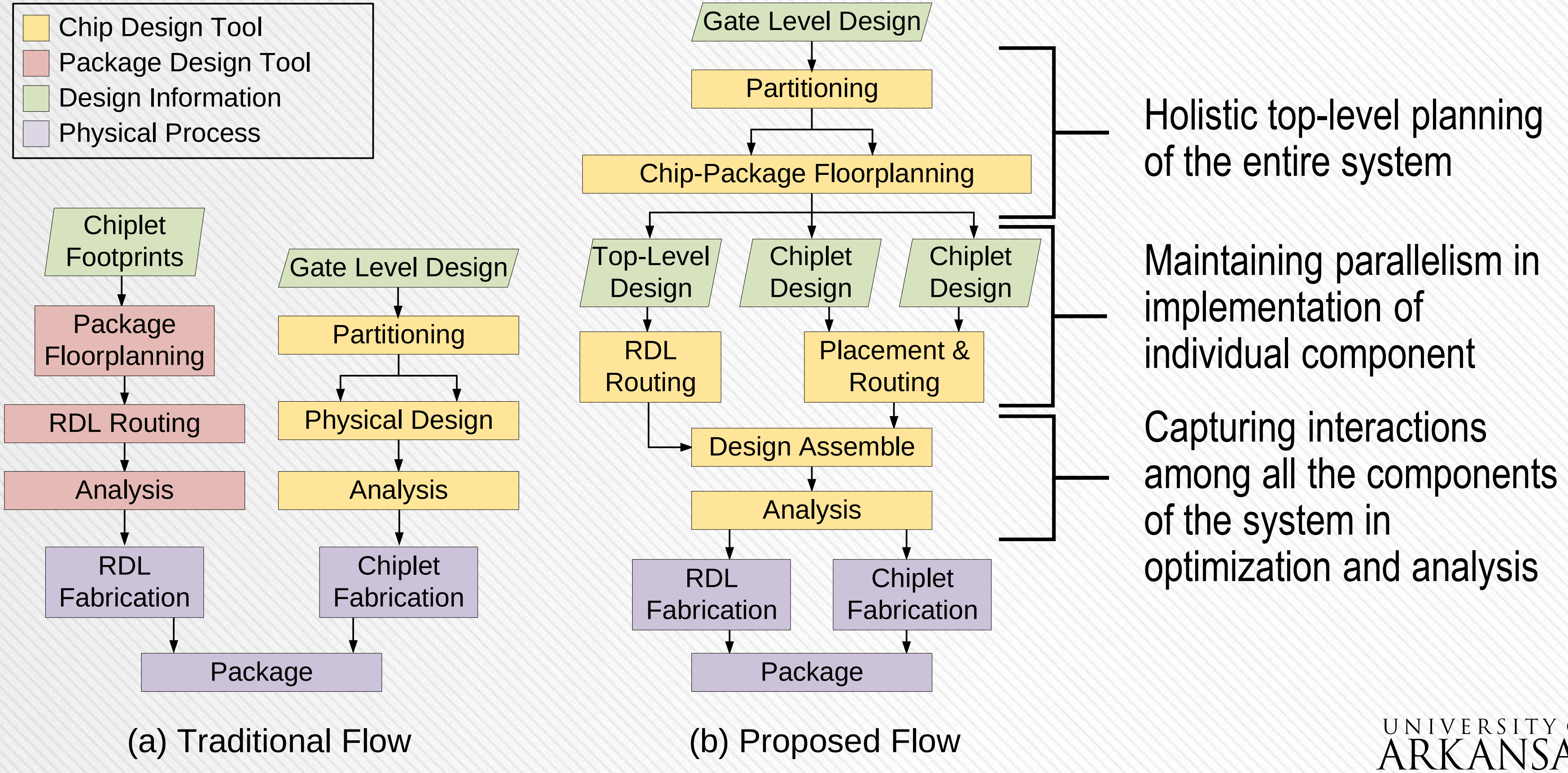
- **Package becomes increasingly critical in post-Moore's Law era**
  - High-density 2D, 2.5D IC, 3D Mem, 3D Sensor, 3D IC, Monolithic 3D IC…
  - Package layers are getting closer and more similar to chip BEOL
- **However, Chip and Package are still two different worlds:**
  - Separate designs by different groups then combined together
  - No existing standard flow that designs 2.5D systems considering chiplets and package interactions during optimization and analysis
  - Optimization goals for 2.5D system are different: inductance, signal integrity, thermal, reliability, cost, turnaround time, flexibility, etc
  - Interactions between the package and chiplets are significant and needs careful consideration in analysis and optimization steps.
- **Objective**
  - Combine chip and package into a single design environment
  - Optimize the entire 2.5D system with detailed package layouts

Traditional Flow vs. Proposed Flow

- **Traditional die-by-die design flow can achieve the shortest possible 2.5D system design time using off-the-shelf chiplets.**
  - Cannot ensure the maximum performance and highest reliability
  - Pin-dominate nature requires both chip and package characteristics
- **Need for a cross-boundary package-aware design strategy:**
  - Pin-dominate nature requires both chip and package characteristics
  - Timing optimization needs to be accounted in the architecture level
  - Partitioner needs to be aware of the delay introduced by redistribution layers (RDLs) with detailed parasitic extraction
  - Package planning tool may need to modify chiplet pin arrangement to optimize RDL routing
  - Chiplet timing optimization steps need to be aware of package wires.
  - The analysis tool needs to consider the chiplets and package interactions altogether.

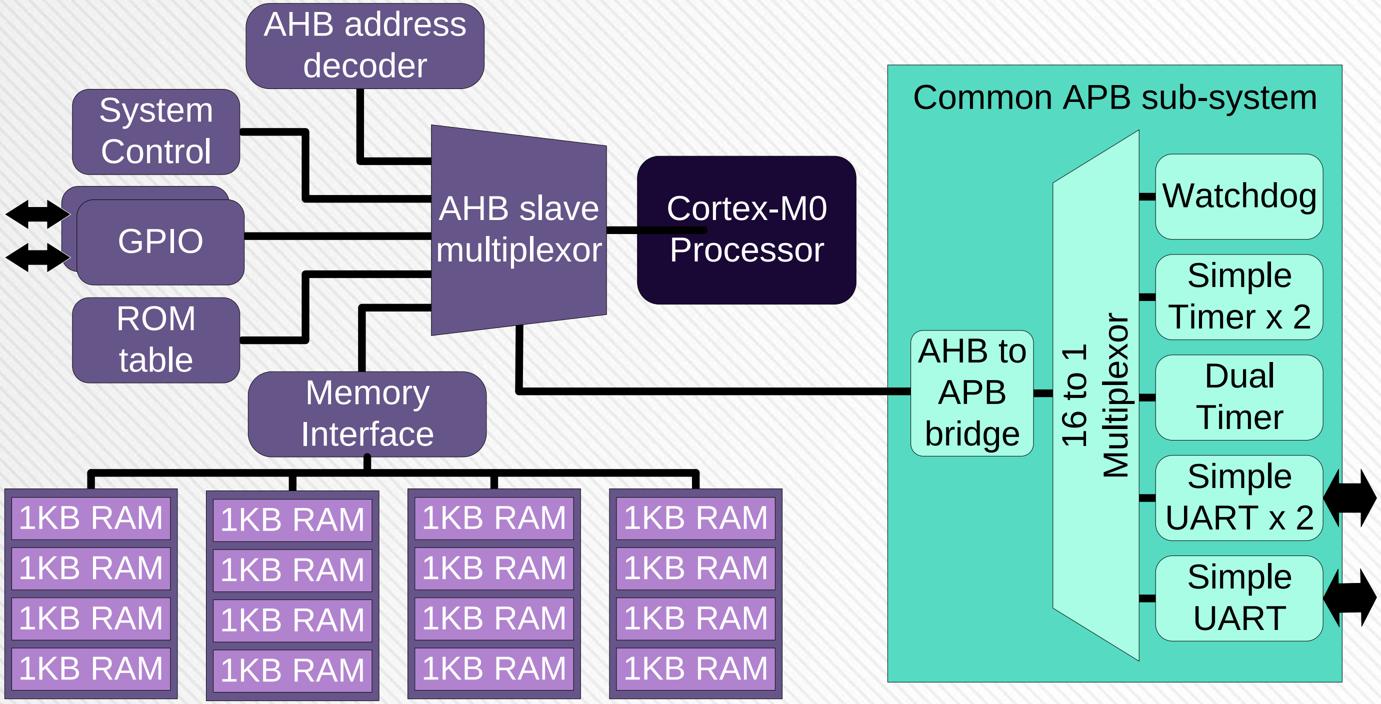
Traditional vs. Our Holistic Flow

- We incorporate the missing necessary interactions between package and chiplets during design, optimization and analysis steps.



Example Design

- System architecture of proof-of-concept design
  - Microcontroller system based on ARM Cortex-M0 core
  - 16KB RAM with some common peripheral devices

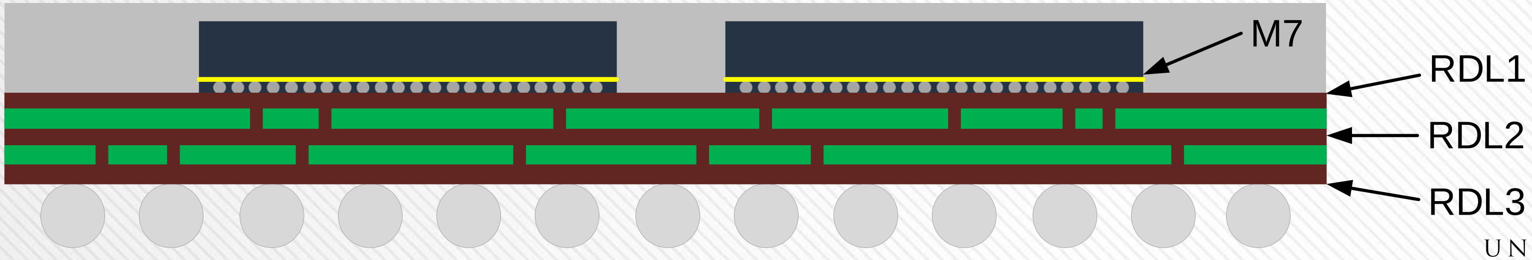


System architecture of example design

Technology Settings

- We use Nangate45nm as our PDK
  - M1-M7 used for chiplet routing
- We modify the top three layers to include 2.5D package RDLs
  - Dimensions are similar to the TSMC 2.5D InFO technology

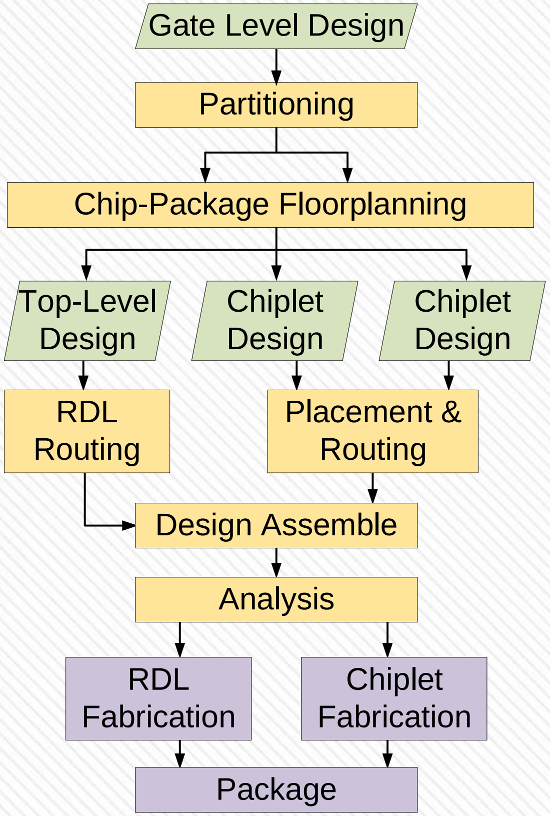
	M6	via6	M7	via7	RDL1	viar1	RDL2	viar2	RDL3
Height	2.28	3.08	3.9	7.5	12.5	17.5	22.5	27.5	32.5
Thickness	0.8	0.82	3.6	5	5	5	5	5	5
Width	0.4	0.4	2	5	10	10	10	10	10
Spacing	0.4	0.44	2	10	10	20	10	20	10



UNIVERSITY OF ARKANSAS

Overall Flow

- Our flow consists of partitioning, top level planning, individual implementation of components, design assembly and analysis.
  - Gate-level netlist is generated by synthesis tool.
  - Entire system is partitioned into chiplets taking into account the impacts of RDLs.
  - Chip-package planning determines the relative locations of chiplets on the package and chiplet pin arrangements. Chiplet floorplanning can also be performed here.
  - Chiplets and package are implemented independently but with constraints propagated from top level to include system-level timing
  - Everything is assembled together for overall optimization and verification



UNIVERSITY OF ARKANSAS

Partition Schemes

- **The partitioner needs to account for package RDL wires while exploring solutions.**
- **We partition the example system into two chiplets.**
  - **Balanced Partition:** Minimum cut size keeping a good area balance. we use **hMetis** and **FLARE** partitioning algorithms as the baseline
  - **Memory vs. Logic Partition:** Based on the natural boundary between memory and logic. We keep all the standard logic cells in one partition and all the memory macros in the other partition.
  - **Architecture-Aware Partition:** In this scheme, we try to utilize the knowledge of the system architecture to come up with a reasonable partition, core-system with memory extension.

Partition Results

- **Despite having little scope of automation, we picked the Architecture Aware Partition scheme for our design, because**
  - Chiplet pins can be accommodated within the chiplet area.
  - This scheme illustrates the drop-in design approach enabled by 2.5D integration technology.

Partition Scheme	hMetis	FLARE	Mem/Logic	Arch-Aware
Max Frequency	300 MHz	300 MHz	333 MHz	300 MHz
Power	6.19 mW	6.17 mW	6.73 mW	6.13 mW
No. of Buffers	2,152	1,907	2,383	2,521
Cell Area (μm^2)	274,450	273,944	275,726	275,902
Area Balance	49.4/50.6	49.8/50.2	89.7/10.3	55.2/44.8
Pin Count	257/313	374/370	191/228	141/110

UNIVERSITY OF ARKANSAS

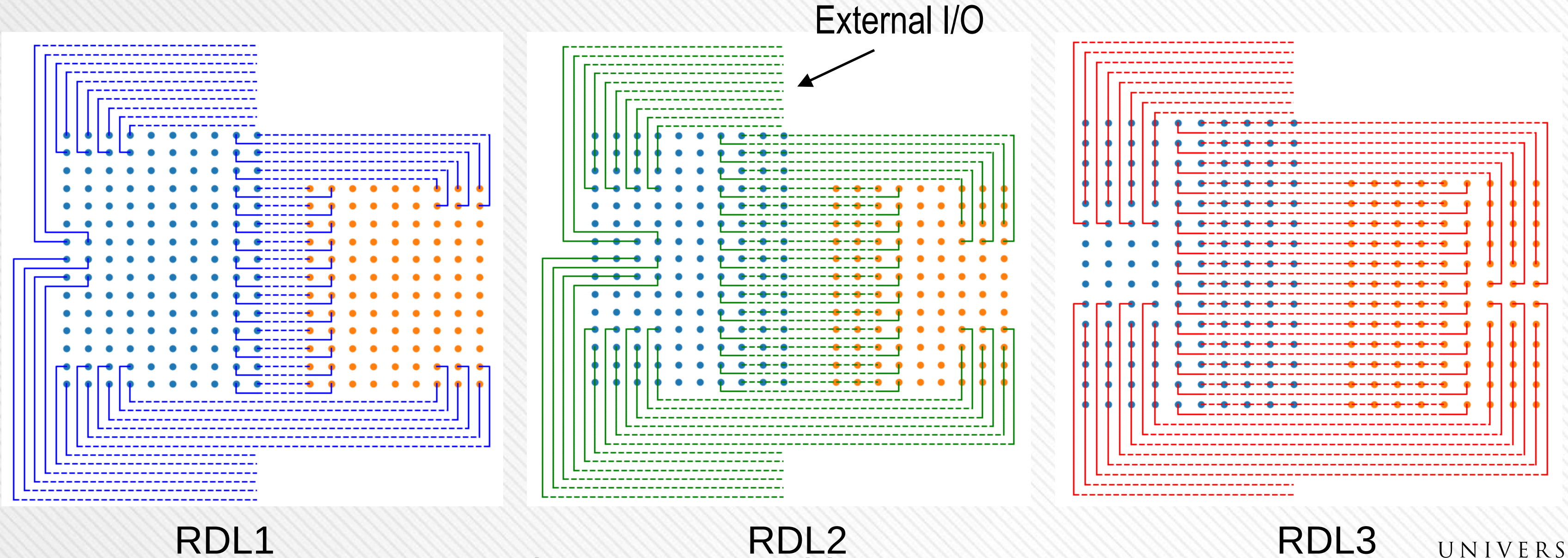
Chip-Package Co-Planning

- **Chip-package planning is performed in two steps**
  - Determine the chiplet dimensions and pin arrangement. Pin arrangement information include pin grid size (rows x cols), pin pitch, pin dimensions, etc.
  - Determine the RDL rouing, chiplet pin connectivity, chiplet pin signal assignment and optionally chiplet floorplanning.
- **The first step is decided from system specifications and floorplaning**
- **For the second step, we have an RDL planning tool that implements a planning strategy and performs following tasks.**
  - Generates RDL routing and routing script for routing tool
  - Generates chiplet pin connectivity, performs signal assignments to the chiplet pins

UNIVERSITY OF ARKANSAS

RDL Routing Strategy

- To minimize long wires and detours on RDLs, we are using following strategies.
  - We don't assign signals to chiplet pins before routing.
  - We route the pins first, and then assign signals based on the routing. This way we have more control and can achieve a very regular routing.
  - Use as many straight wires as possible to connect the chiplet pins.

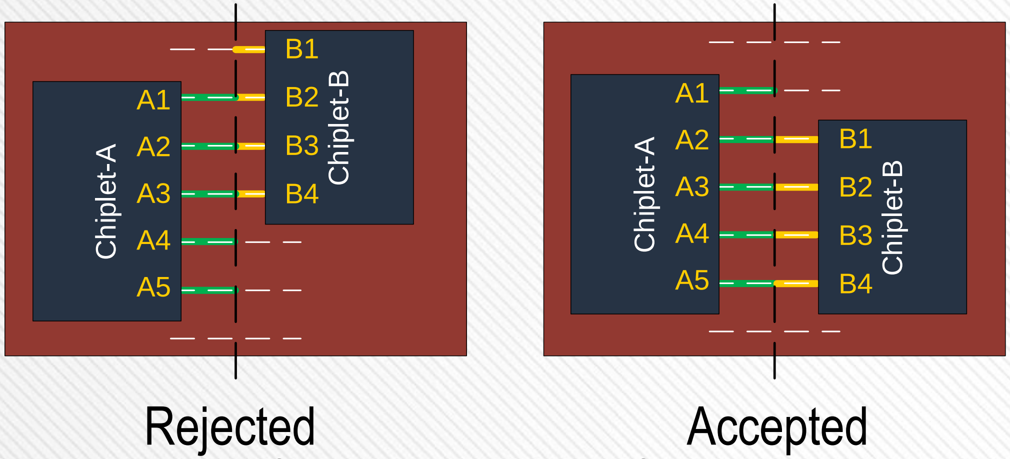


Routing Generated by our program

UNIVERSITY OF ARKANSAS

Chip-Package Floorplanning

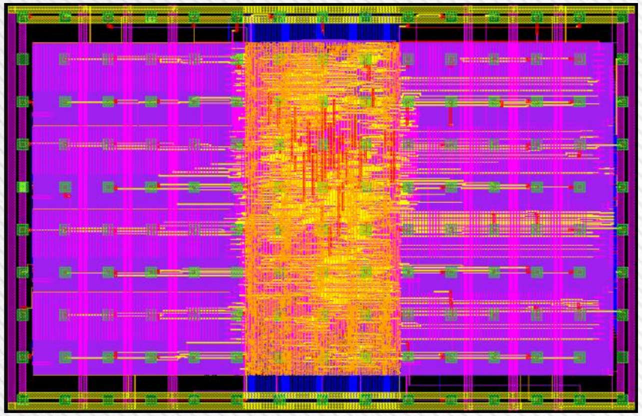
- **The relative locations of the chiplets are determined during RDL routing.**
  - Chiplet floorplanning is determined based on pin alignment
  - We perform only point-to-point connection since this is the most commonly used wire connection on the package level. Shared bus is generally for low-speed communications
  - This straightforward strategy is the first step. Steiner routing can be used to improve performance and handle multi-point connection



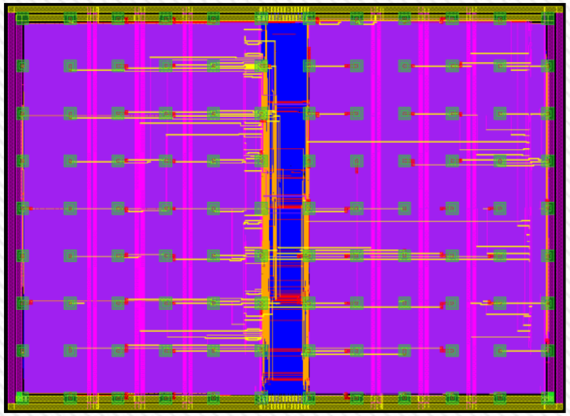
Determination of relative location of the chiplets on RDL

Placement & Routing

- ❑ **After top level planning, chiplets and package are implemented independently with constraints propagated from top-level**
  - Top level design is hierarchically splited like 2D partitioning.
  - Chiplet floorplan may change as required, only the pin arrangement needs to be the same as fixed by top level planning.
  - Chiplet implementation is the same as conventional 2D chip that includes power planning, placement, time design, routing and post routing optimizations.



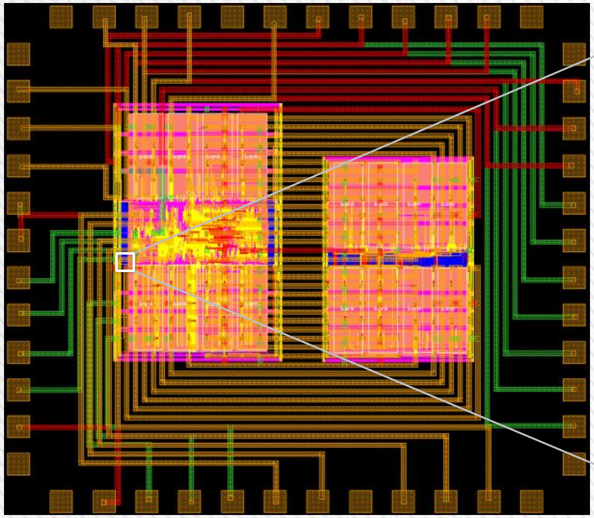
(a) Core System Chiplet



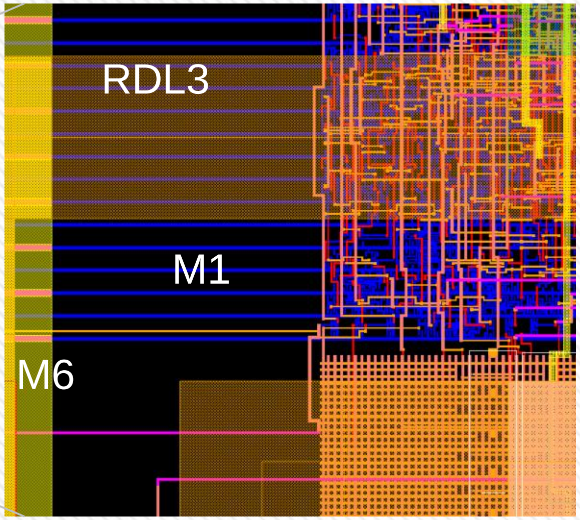
(b) Extended Memory Chiplet

Design Assemble

- After chiplet implementation, the entire system is assembled into a hierarchical design layout
  - As the design environment has everything together, some incremental optimizations can be performed to improve overall system performance.
  - The analysis and optimization tools have all the information needed to account for the impacts of RDLs on chiplet design.



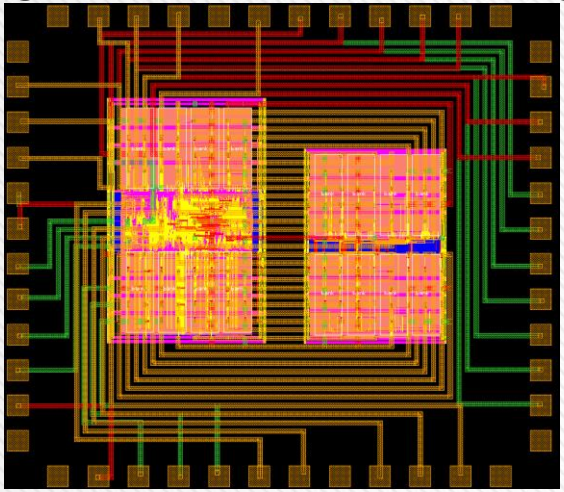
(a) Ext. Memory 2.5D System



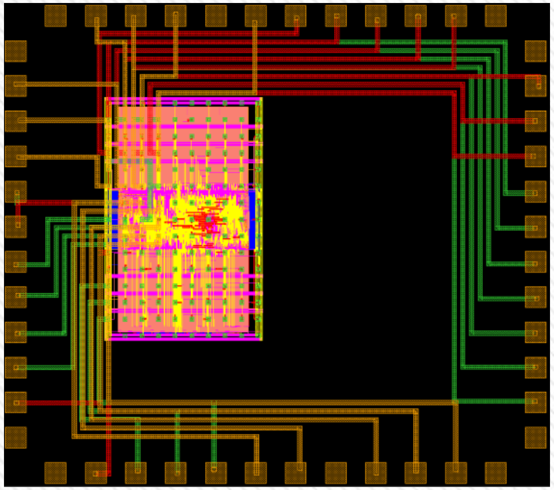
(b) Zoom-in Shot

Drop-In Design

- **We design two versions of the 2.5D system using "drop-in" design technique enabled by 2.5D integration**
  - Provides design flexibility and product binning in no time
  - The core-chiplet contains all the logic blocks and 8KB of memory and can be used without the memory extension chiplet.
  - The memory chiplet contains extra 8KB of memory.
  - No change required in chiplet designs, the same package design can be used.



16KB System, suitable for memory extensive applications

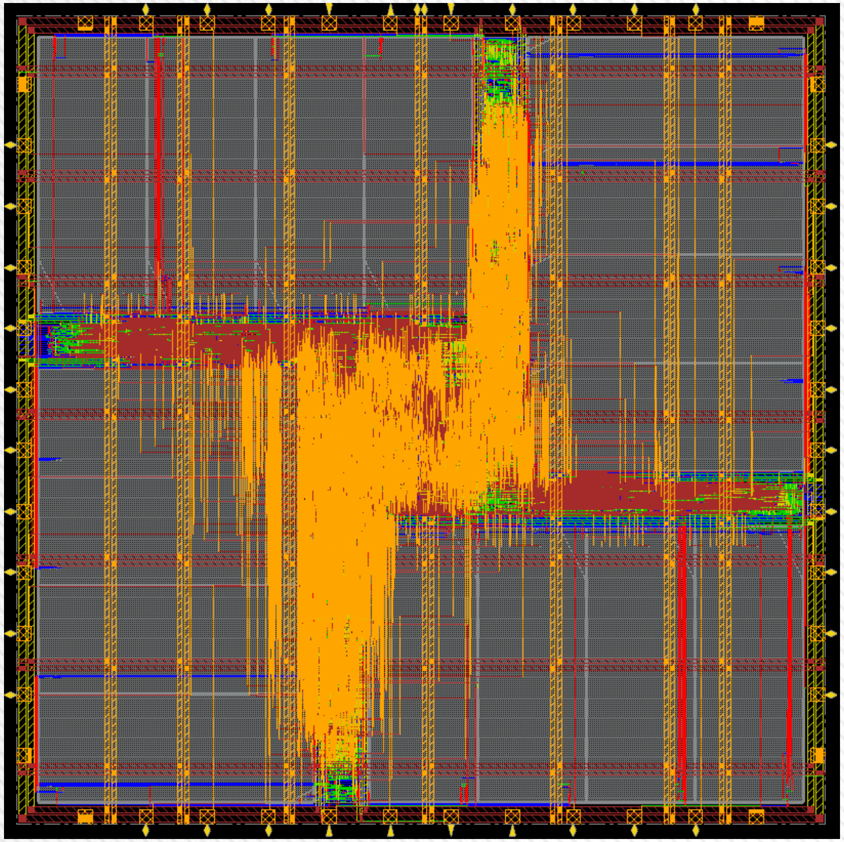


8KB System, a cheaper solution for applications requiring small memory

UNIVERSITY OF ARKANSAS

Reference 2D Design

- For comparative study, we implemented the same system as monolithic 2D die using conventional 2D chip design flow
  - Only with 16KB mem flavor



❑ Chiplet-Package coupling capacitance

- After design assemble, we exported the entire system in GDS format for parasitic extraction.
- The columns for RDL1, RDL2, and RDL3 show the coupling capacitance between package layers and chiplet layers (in fF).
- The coupling of package layers with M7 is low because of a smaller number of wires on M7. However, there exists significant coupling with the wires on M6, which is captured in the parasitic extraction process
- Also, M7 and RDL1 are extracted with considerations from the other side

	M1-M5	M6	M7	RDL1	RDL2	RDL3
M1-M5	5187	479.1	22.52	58.16	9.889	7.547
M6	479.1	533.7	84.89	101.1	11.62	10.85
M7	22.52	84.89	26.68	14.84	1.739	1.663
rdl1	58.16	101.1	14.84	297.1	1009	41.49
rdl2	9.889	11.62	1.739	1009	297.4	1076
rdl3	7.547	10.85	1.663	41.49	1076	513.1

2D vs 2.5D Systems Comparison

Die-level Chiplet analysis results

- Analysis captures the impact of RDLs on system performance.
- For the 2.5D system the highest system frequency we achieve is 245 MHz which is much worse, as expected, compared to the maximum frequency (333 MHz) of the 2D implementation.
- To achieve this performance in 2.5D system, a large number of buffers are needed which is reflected in the cell count
  - No buffer can be used on the package level -> requires different timining optimization strategy

Chip Design	2D Monolithic	Core Chiplet	Ext. Mem Chiplet
Standard Cells#	35904	51733	11531
Total Chip WL	412.990 mm	350.898 mm	40.143 mm
Die Size	550x550μm^2	390x590μm^2	350x470μm^2
Frequency	333MHz	245MHz	
Chip Power	10.6mW	7.751 mW	0.194 mW

UNIVERSITY OF ARKANSAS

Drop-In Designs Comparison

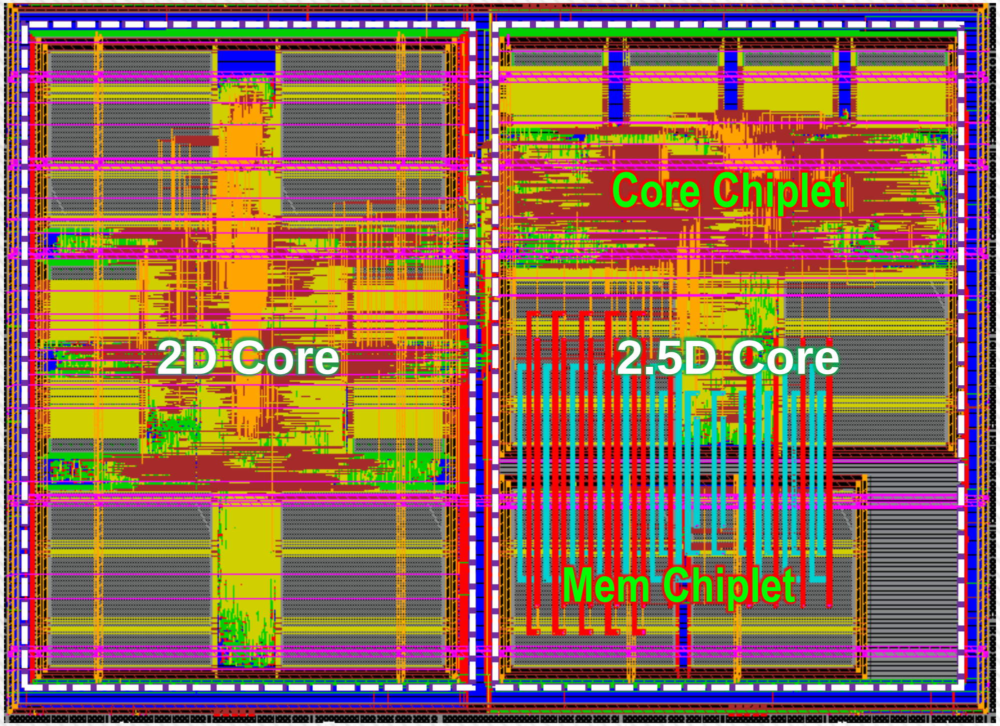
- **Performance trade-off between the two versions of 2.5D systems is evident from analysis**
  - **The Chip-Package coupling capacitance is larger for the extended memory system because of more package wires**
    - Package RC delay is successfully captured in full-system analysis
  - **The critical timing path for the extended system is between the core and memory chiplets.**
  - **In the absence of the extra memory chiplet, we could achieve a higher system frequency for the Core-Only system**

System Design	Core-Chiplet Only	Core-Mem Chiplets
Chip-Package Cap	120.7864 fF	217.4089 fF
Max Frequency	300 MHz	245 MHz
System Power	9.578 mW	8.26 mW
Package wirelength	35.41 mm	94.027 mm
Package Size	1.3mm x 1.15mm	

Latest Update

❑ Tape-out Design with TSMC 65nm using our holistic flow

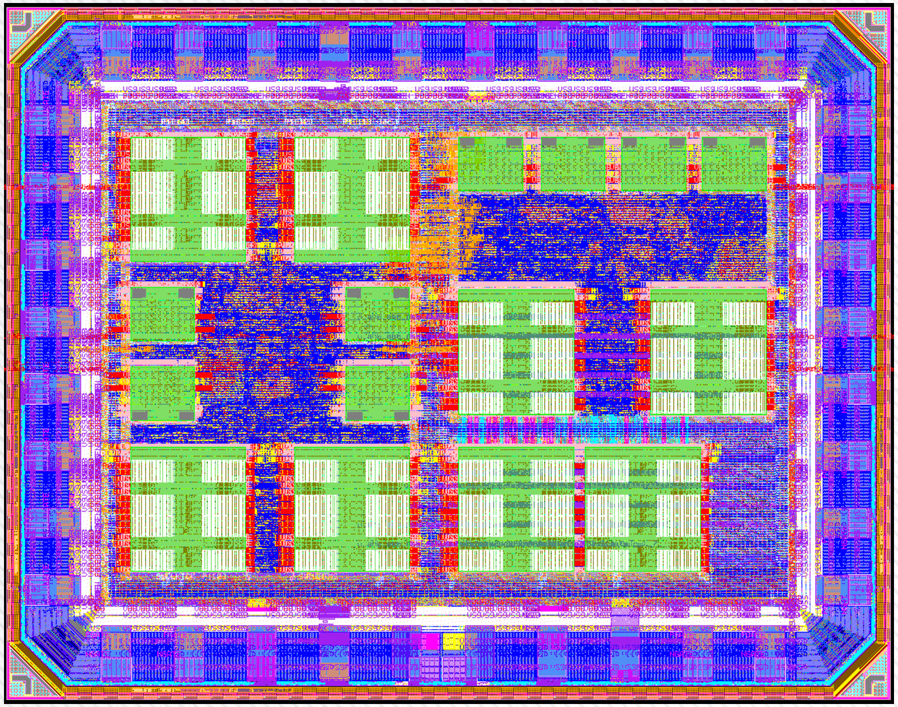
- Power Distribution Network is designed at the top level.
- Placement, CTS, routing, etc. are performed on the top level as it includes the standard cells of the pin mux module.
- After finishing all the steps, all of the designs are assembled at the top level.



Tape-out Design with TSMC 65nm

❑ Everything together

- We separately created a bondpad GDS wrapper for the design as the I/O pad cells don’t have the bondpads in them.
- A sealring is created from the reference sealring GDS provided by the foundry.
- Dummy metal/poly shapes are generated using foundry provided scripts and then merged with our design using Calibre DRV file merge command.



UNIVERSITY OF ARKANSAS

Conclusions & Future Work

Conclusions

- Chiplet-Package interactions need to be considered early in design
- Our flow effectively captures the impact of RDLs in optimization and analysis steps.
- It incorporates necessary interactions between package and chiplet designs for holistic planning and optimization.
- The flow is suitable for homogeneous designs with existing commercial chip design tools.

Future Works

- Chiplet-Package inductance impact on PPA and noise
- More sophisticated algorithms with 45-degree routing, multi-point connection, diff-pair routing, PG ground planning and filling is needed
- New tools/techniques based on in-context design strategy need to be developed to support heterogeneous designs.

UNIVERSITY OF ARKANSAS

http://csce.uark.edu +1 (479) 575-6043 yrpeng@uark.edu